REMARKS

Claims 1, 15, 30, 33 and 37 are amended herein.

Claims 2, 11-13, 16, 25-27, 31 and 34 are canceled.

Claims 1, 3-10, 14-15, 17-24, 28-30, 32-33, and 35-37 are now pending.

Response to Rejection Under 35 USC §103(a)

Claims 1, 2, 4, 6, 9-10, 14-16, 18, 20, 23, 24, and 28-37 are rejected under 35 U.S.C. 103(a) as being unpatentable over van den Akker (US 6,415,250) in view of Bracewell et al (US PGPub 2006/0041685) and in further view of de Campos (US6,272,456). Applicants respectfully traverse.

As amended, claim 1 recites a system, comprising:

> a storage system adapted to store a set of language classes, which each identify a language and a character set encoding, and further adapted to store a plurality of training documents;
> an attribute modeler adapted to train an attribute model by evaluating occurrences of one or more document properties within the training documents and, for each language class, calculating a probability for the document properties set conditioned on the occurrence of the language class, the trained attribute model stored in the storage;
> a text modeler adapted to train a text model by evaluating byte occurrences within the training documents and, for each language class, calculating a probability for the byte occurrences conditioned on the occurrence of the language class, the trained text model stored in the storage; and
> a training engine adapted to calculate an overall probability for ones of the set of language classes by evaluating the probability for the document properties set based on the attribute model and the probability for the byte occurrences based on the text model.

. These features provide a system for identifying language attributes through probabilistic analysis using two probabilistic models and a training engine on a plurality of training documents. For ones in the set of language classes, an attribute model calculates a probability for the document properties set within the training documents conditioned on the occurrence of

the language class, and the text model calculates a probability for the byte occurrences within the training documents conditioned on the occurrence of the language class. The training engine calculates an overall probability for language classes by evaluating the probability for the document properties set based on the attribute model and the text model.

van den Akker does not teach or suggest at least these features. As admitted by the Examiner, van den Akker does not disclose "calculating a probability for the byte occurrences conditioned on the occurrence of the language class" (see second element of claim 1). van den Akker's language identification system attempts to identify the language of a document based on an analysis of fixed length word suffixes sampled from input text. See, for example, van den Akker, Abstract or column 10, lines 40-45. van den Akker's probability represents a relative likelihood that a given text is the associated language due to the presence of the associated suffix in the text. See van den Akker, column 3, lines 15-40. In other words, van den Akker identifies language as function of word suffixes.

In addition, as admitted by the Examiner, van den Akker does not disclose training "an attribute model" "by evaluating occurrences of one or more document properties within the training documents and, for each language class, calculating a probability for the document properties set conditioned on the occurrence of the language class" (see third element of claim 1) Instead, at least one version of van den Akker's language identification system eliminates character set encoding information entirely by converting all documents, regardless of language encoding, into ANSI text before identifying the language of the document. See, for example, van den Akker, column 11, lines 33-39. As such, van den Akker does not use the character set encoding as recited in the independent claims . Furthermore, since van der Akker does not disclose either the claimed attribute model or the claimed text model, van den Akker does not disclose calculating "an overall probability for ones of the language classes by evaluating the probability for the document properties based on the attribute model and the probability for the byte

occurrences based on the text model." (see fourth element of claim 1). Instead, van den Akker's language identification engine determines for each language class an arithmetic sum of the relative probabilities for all the suffixes which appear in the text. See van den Akker, column 3, lines 40-45. Thus, van der Akken fails to disclose at least the second, third, and fourth elements of claim 1 for the reasons discussed above.

Bracewell does not remedy the deficiencies of van den Akker. Bracewell discloses a system for rendering data that is not natively comparable for reviewing on web browsers. Bracewell's template relies on the HTTP request to identify the language type. See Bracewell, ¶14. More specifically, Bracewell assumes that the language of a document is explicitly identified in the HTTP request, because only then can Bracewell select the correct template. As such, Bracewell does not disclose identifying language attributes through a probabilistic analysis of the document properties by an attribute model as claimed.

de Campos does not remedy the deficiencies of van den Akker and Bracewell. de Campos discloses using a combination of different sized n-gram language profiles to identify the language of a document. See de Campos, column 11, lines 49-53. de Campos teaches identifying a window of letters by evaluating the frequency parameters of the matched reference letter sequences in each language profile. See de Campos, column 10, lines 46-65. As such, de Campos does not disclose or teaching using an attribute model and a text model to identify language attributes through probabilistic analysis as claimed.

The Examiner's proposed modification of van den Akker results in relying on two models not disclosed or suggested in van den Akker, where van den Akker only determines for each language class an arithmetic sum of the relative probabilities for all the suffixes which appear in the text.

Claim 1 is amended to include the limitations of claim 2. The examiner's rejection of claim 2 on page 4 of the Office Action only discussed van der Akker. If the Examiner persists in this rejection, applicants request that the examiner explain his rejection is light of the other cited references.

Independent claim 15 as amended is similar to claim 1 and patentably distinguishes over the cited documents for at least the same reasons as discussed above in connection with claim 1.

Independent claims 30, 33, and 37 for example, recite a system, method, and apparatus including, for example, calculating an overall probability for ones of the set of language classes by evaluating the probability for the top level domain and character set encoding based on the attribute model and the probability for the byte occurrences based on the text model. Claims 30, 33, and 37 are patentably distinct for at least the same reasons as claims 1 and 15 discussed above. Moreover, the Examiner admits that van den Akker does not disclose calculating an overall probability for ones of the set of language classes by evaluating the probability for the top level domain and character set encoding based on the attribute model and the probability for the byte occurrences based on the text model. See, for example, van den Akker, column 11, lines 33-39.

Claims 3, 5, 17 and 19 are rejected under 35 U.S.C. 103(a) as being unpatentable over van den Akker in view of Bracewell et al. applied to claims 1, 2, 4, 6, 9-16, 18, 20, 23-37 above, and in further view of Elworthy (US 6,125, 362). Elworthy does not remedy the deficiencies of van den Akker, Bracewell and de Campos. Elworthy teaches a Bayesian probabilistic formula to classify documents based on an assumed language classification and the probability of an element that is the text of the document. Specifically, in Elworthy's Equation 1, p(t) is the probability of a text token given a language classification, and thus not the probability of the document properties and actual text as claimed. Furthermore, Elworthy does not teach that a language class is defined by a language and a character set encoding as claimed.

For at least the reasons above, Applicants submit that claims 1, 15, 30, 33, and 37 are patentable over the cited references. Claims 3-10, 14, 17-24, 28-29, 32, and 35-36 either directly or indirectly depend from claims 1, 15, 30 and 33. These dependent claims also recite additional features not disclosed by the cited references. Thus, Applicants submit claims 3-10, 14, 17-24, 28-29, 32, and 35-36 are patentably distinguishable over the cited references

In sum, Applicants submit that the pending claims are patentably distinguishable over the cited references. Therefore, Applicants request reconsideration of the basis for the rejections to these claims and request allowance of them. If the Examiner is in need of further information, he is invited to contact the undersigned attorney at the telephone number provided below.

Respectfully submitted,
Alex Franz et al.

Dated: ___November 7, 2007____

By: __Robert R. Sachs/___
Robert R. Sachs, Reg. No. 42,120
Attorney for Applicants
Fenwick & West LLP
Silicon Valley Center
801 California Street
Mountain View, CA 94041
Tel.: (415) 875-2410
Fax: (415) 281-1350